University of Rochester

DSC 275/475: Time-Series Analysis and Forecasting for Data Science

Fall 2019

Location and Time:

Tuesdays and Thursdays 11:05-12:20

CSB 209

Instructor

Prof. Ajay Anand, Wegmans Hall Rm 1203, x63149 (ajay.anand@rochester.edu)
Office hours: Thursday (13:00-14:00) or by appointment (use email).

Dr. Edgar A. Bernal, Wegmans Hall Rm 1209, 585 481 9105 (edgar.bernal@rochester.edu)
Office hours: by appointment (use email).

TAs/Office hours:

Zhou (Joe) Xu zxu17@ur.rochester.edu

Xiaoning (Isaac) Guo xguo24@u.Rochester.edu

Faner Lin flin11@u.Rochester.edu

Course description

Time series analysis is a valuable data analysis technique in a variety of industrial (e.g., prognostics and health management), business (e.g., financial data analysis) and healthcare (e.g., disease progression modeling) applications. Moreover, forecasting in time series is an essential component of predictive analytics. The course will begin with an introduction to practical aspects relevant to

time series data analysis such as data collection, characterization, and preprocessing. Topics covered will include smoothing methods (moving average, exponential smoothing), trend and seasonality in regression models, autocorrelation, AR and ARIMA models, and application of neural network (including deep learning-based) models to time-series data. Students shall work on projects with time-series data sets using modeling tools in Python/R.

Prerequisites: MTH 165 and CSC161/CSC171 or equivalent intro programming coursework, introductory statistics course (STT 212/213, DSC 262) is recommended.

Required Text book:

Introduction to Time Series Analysis and Forecasting (Wiley Series in Probability and Statistics) 2nd Edition by *Douglas C. Montgomery (Author), Cheryl L. Jennings (Author), Murat Kulahci (Author)*

- Series: Wiley Series in Probability and Statistics
- Hardcover: 672 pages
- Publisher: Wiley-Interscience; 2 edition (April 27, 2015)
- Language: English
- ISBN-10: 1118745116
- ISBN-13: 978-1118745113

Deep Learning (Adaptive Computation and Machine Learning series) *by Ian Goodfellow (Author), Yoshua Bengio (Author), Aaron Courville (Author)*

- Series: Adaptive Computation and Machine Learning series
- Hardcover: 800 pages
- Publisher: The MIT Press (November 18, 2016)
- Language: English
- ISBN-10: 0262035618
- ISBN-13: 978-0262035613
- Available https://www.deeplearningbook.org/

Additional Reading:

Practical Time Series Forecasting with R: A Hands-On Guide [2nd Edition] (Practical Analytics) 2nd Edition by Galit Shmueli (Author), Kenneth C. Lichtendahl Jr (Author)

Pattern Recognition and Machine Learning (Information Science and Statistics) by Christopher M. Bishop (Author) (available https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf***)***

Course schedule (chapters refer to the textbook)

Topic	Instructor	Textbook Chapter Reference
Introduction to Time-series Analysis and Forecasting	Anand	(Chap. 1, Montgomery)
Statistics introduction to Time-series Analysis	Anand	(Chap. 2, Montgomery)
Smoothing Methods (MA, Exponential Smoothing)	Anand	(Chap. 4, Montgomery)
Auto-regressive Techniques (AR, ARMA, ARIMA)	Anand	(Chap. 5, Montgomery)
Time-series Forecasting using Machine Learning Techniques	Anand	(Chap. 8, Shmueli) and Notes/slides
Additional Topics (e.g., Spectral analysis, multivariate time-series, Markov chains) *(Tentative)*	Anand	(Chap. 7, Montgomery) and Notes
Basics of Deep Learning	Bernal	(Chap. 5, Goodfellow and Chap. 3-4 Bishop)
Introduction to Neural Networks (optimization, backpropagation, regularization)	Bernal	(Chap. 6-8, Goodfellow and Chap. 5 Bishop)
Vanilla Recurrent Neural Networks	Bernal	(Chap. 10, Goodfellow)
Gated Recurrent Neural Networks	Bernal	(Chap. 10, Goodfellow)
Convolutional Neural Networks	Bernal	(Chap. 9, Goodfellow)
Temporal Convolutional Neural Networks	Bernal	Notes

Course Schedule and Outline:

Midterm Exam: October 10th, 2019

(Will be posted on a Google Doc as a live link with updates)

Due to the nature of the course being dependent on projects from external sponsors the above outline is subject to change. Changes will be published as soon as is reasonable to minimize disruptions, but changes should be expected.

Grading

Grading (total 100%) [Graduate students will have additional problems in select assignments]

- Homework assignments: 20% (3 HWs) – 1st half of course
- Midterm: 25%
- Project #1: 15% (based on material covered until mid-term)
- Project #2: 40% (made of 4 structured modules- each 10%, based on latter-half of course)

Homeworks and Projects are due at 11:59 pm on the due date. **Late points** will be subtracted (2% per hour). Must submit on **Blackboard** online. Only the latest submission is considered.

Graduate students will have additional problems in homeworks and project #1 to receive graduate level credit.

Re-grading deadline is 1 week from the date graded assignment is returned.

Exam policy: No make-up exams will be given. If a student has to miss an exam for an unavoidable circumstance, please discuss with the instructor as soon as possible.

Programming Experience

The HWs and the project require implementing time series algorithms. You are welcome to choose a programming language of your choice for the HWs and Project #1 (e.g. Python, R, Matlab). Project #2 is based on using Deep-learning libraries in Python. Instructor will lead hands-on programming sessions during class for Project #2 with follow on assignments to be completed outside class.

Project #2 Overview:

The goal of the project is to develop statistically informed model for stock trading in order to maximize profit. To this end, daily stock prices (in the form of opening and closing price, as well as low and high session) for 20 companies across a 10-year period will be provided. The efficacy of the algorithms will be evaluated based on two criteria: (i) the resulting prediction MSE across a six-month period relative to held-out data, and (ii) the resulting profit margins based on a set of hypothetical operations across the same six-month period. Algorithms based on RNNs, gated RNNs and temporal convolutional networks will be developed, with free range with regards to the choice of architecture hyper parameters and data manipulation, as long as causality constraints are not violated.

Academic Honesty Policy

Academic Honesty Policy: The University of Rochester academic honesty policy applies to all assignments and exams for this class. The full-text of the academic honesty policy can be found at: http://www.rochester.edu/College/CCAS/AdviserHandbook/AcadHonesty.html In addition to the general guidelines mentioned in the above policy, for this course I require that: In examinations, you must work individually with no communication with others and use only materials/tools that have been explicitly allowed. For homework, you may discuss problems with your colleagues but final solutions need to be worked out, written and submitted individually. Any external material used should be clearly cited. In your own writings (example term papers, homework solutions, proposals etc), no more than one or two sentences may be used verbatim from any source. READ THESE INSTRUCTIONS CAREFULLY! If any aspect of the academic honesty policy and guidelines for this course are unclear, please ask me for clarifications. Lack of awareness or understanding of this policy will not be an acceptable excuse or defense against disciplinary action.

Credit Hour Policy:

This course follows the College credit hour policy for four-credit courses. This course meets two times weekly for three academic hours per week. The course requires students to complete homework, mini-project, team project, and a mid-term exam. The students are expected to complete supplementary work averaging 8 hours per week. The team project would be done in pairs of two. The student activities on the homework and projects would include developing algorithms, writing code, explaining concepts learned in class, and preparing documentation.